W.J. Usery Workplace Research Group Paper Series

Working Paper 2009-6-1
June 2009

Inequality and Volatility Moderation in Russia: Evidence from Micro-Level

Yuriy Gorodnichenko
University of California, Berkeley

Klara Sabirianova Peter
Georgia State University

Dmitriy Stolyarov
University of Michigan

Georgia State University
Andrew Young School of Policy Studies

Inequality and Volatility Moderation in Russia: Evidence from Micro-Level Panel Data on Consumption and Income

June, 2009

Yuriy Gorodnichenko
UC Berkeley
NBER and IZA
ygorodni@econ.berkeley.edu

Klara Sabirianova Peter
Georgia State University
CEPR and IZA
kpeter@gsu.edu

Dmitriy Stolyarov
University of Michigan
stolyar@umich.edu

Abstract

We construct key household and individual economic variables using a panel micro data set from the Russia Longitudinal Monitoring Survey (RLMS) for 1994-2005. We analyze cross-sectional income and consumption inequality and find that inequality decreased during the 2000-2005 economic recovery. The decrease appears to be driven by falling volatility of transitory income shocks. The response of consumption to permanent and transitory income shocks becomes weaker later in the sample, consistent with greater self-insurance against permanent shocks and greater smoothing of transitory shocks. Comparisons of RLMS data with official macroeconomic statistics reveal that national accounts may underestimate the extent of unofficial economic activity, and that the official consumer price index may overstate inflation and be prone to quality bias.

Keywords: inequality, income, consumption, transition, Russia.

JEL Classification: E20, J30, I30, O15, P20

Acknowledgement: We thank Steven Stillman and conference and seminar participants at IZA, the New Economic School, and the PIER IGIER Fifth International Conference for useful comments.

1. Introduction

Modern macroeconomists are increasingly relying on the analysis of environments with heterogeneous agents. Many macroeconomic questions can only be asked (and answered) in the context of multi-agent environments. These richer macroeconomic models require a correspondingly rich set of empirical facts that come from micro data and incorporate information on distributions in addition to the usual aggregates. The goal of this paper is to provide a comprehensive set of cross-sectional and time series stylized facts for the Russian economy and a systematic study of multiple dimensions of inequality.

Since the late 1980s, Russian economy has been subject to substantial macroeconomic volatility, with a long phase of severe output contraction, periods of high and variable inflation, and a subsequent period of recovery. At the same time, Russia has tremendous regional diversity. The combination of these factors presents unique opportunities for studying both cross-sectional and time-varying dimensions of inequality. Fortunately, high quality data are available to explore these opportunities: a large, nationally representative panel study of Russian households that incorporates economic variables, the Russia Longitudinal Monitoring Survey (RLMS).

This paper includes multiple dimensions of inequality, with particular focus on consumption and income. We construct the key variables describing the economic behavior of Russian households and individuals and analyze their cross-sectional dispersion and time series patterns. Specifically, we create time-varying distributions of individual earnings and labor supply, as well as household-level income, expenditure, and consumption.

We would like to highlight two main results. First, almost all measures of cross-sectional inequality in income and consumption started falling during 2000-2005, after staying relatively high during 1994-1998. Second, the measured fall in inequality is mostly due to the moderation of the transitory shocks to household income and consumption.

The recent period of falling inequality was preceded by an initial rise in the early 1990s that accompanied Russia's transition from a centrally planned to market economy (e.g., Commander *et al* 1999, Galbraith *et al* 2004). However, the level of inequality at the end of our sample is still higher than it was during the socialist era. Interestingly, poor households do not appear to fall behind during the economic recovery – the lower tail of the expenditure distribution does not diverge from the middle as the economy expands. The latest level of inequality that we find is typical for a middle income country. For example, the Gini coefficient in 2005 was about 0.38-0.40, which is just slightly above the mean value of Gini coefficients for after-tax household income and consumption from upper middle income countries.[1,2]

Some features that set the Russian economy apart from more developed countries turn out to be important for the analysis of inequality. One such feature is home production of food. Our results indicate that home-grown food has a large equalizing effect on income and consumption. The effect is large, because poorer rural households are also the ones that grow a lot of food for own consumption. Another unique feature of the Russian economy is its geographic diversity. Accounting for regional differences in the cost of living (that vary by a factor of 2.7 in Russia) is shown to have a sizeable equalizing effect. Other important features of the Russian transition, such as underreporting of income, wage payment delays, irregularities in government transfer payments, and forced in-kind substitutes in lieu of wage payments also explain some of the inequality trends.

The comparison of income and expenditure inequality reveals further differences from developed economies. In developed economies, expenditures are usually distributed more equally than income, which is attributed to consumption smoothing possibilities. This turns out not to be the case for Russia, where expenditure inequality is almost as high as income inequality. We argue that the relatively high expenditure inequality reflected peculiar patterns of consumption smoothing during the downturn. Households facing irregular wage and transfer

---

[1] Our results on inequality levels have to be taken in the context of our sample. We think that the RLMS, like most household surveys, may under-represent the very rich individuals who own capital assets in Russia. This is evident from the negligible financial asset holdings of most RLMS respondents. The studies that attempt to adjust for super-rich typically document much higher levels of inequality. For example, Guriev and Rachinsky (2006) find that the income Gini coefficient for the city of Moscow is 0.625, and Aivazian and Kolenikov (2001) report a Gini coefficient of 0.55-0.57 based on parametric estimation of the uncensored expenditure distribution. We find some evidence that suggests divergence between the super-rich and the rest of the population in 2003-2005 (see Section 2 for further discussion).

[2] The comparisons are made using the Inequality Database of the World Institute for Development Economics Research.

payments, high inflation, and undeveloped financial markets used less conventional mechanisms such as food storage to smooth consumption. Food inventories were built up when income was received to insure against irregular wage payments.

We further look at the inequality dynamics between groups in our sample. We find the comparison of economic experience between urban and rural populations is particularly interesting. The rural population has a more restricted choice of jobs, which limits occupational mobility during transition. In addition, the workers with highest earnings potential might have migrated to cities. However, we do not find evidence that income and consumption of the rural population fell behind. The rural group did not seem to do relatively worse during the downturn, although during the recovery the rural population exhibited a slower growth rate in consumption of durables.

More broadly, we have found almost no evidence of convergence or divergence between groups based on observables, such as education, location, household composition, and age. The reduction in inequality during economic recovery resulted mostly from the moderation in the residual volatility of income and consumption growth.

We examine the reasons for the observed fall in residual income volatility by exploiting the panel dimensions of the data. In particular, we decompose the income process into permanent and transitory components and estimate their effect on consumption. We document that the fall in residual income volatility is mostly due to a fall in the variance of transitory income shocks.[3] Over time, consumption response to both permanent and transitory income components becomes weaker. This is consistent with better insurance against income shocks and hence better consumption smoothing later in the mid 2000s.

Apart from the analysis of inequality trends, we examine the trends in the levels of income and expenditure. The time trends show a 40 percent drop in real per-capita expenditure and a 50 percent drop in real hourly wages during 1994-1998. Recent literature has argued that the drop in Russian real output during the transition has been overstated due to exaggeration of the Soviet output and mismeasurement of the unofficial economy in the 1990s (Schleifer and

---

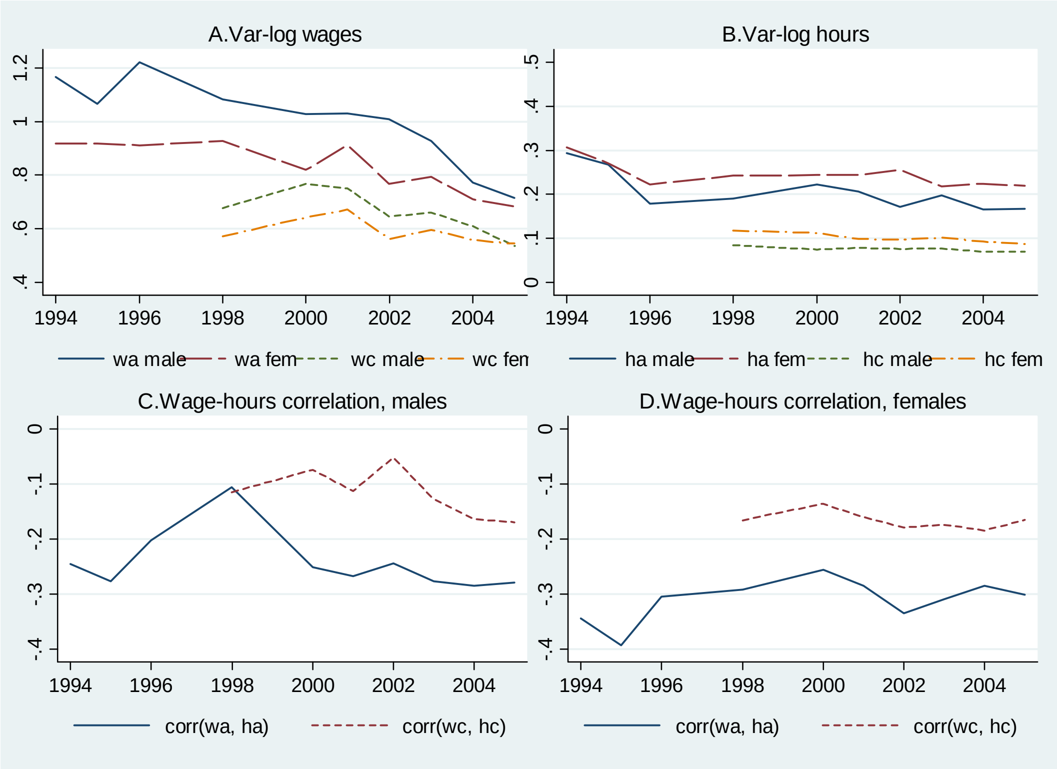
[3] Stillman (2001) finds that RLMS expenditures respond strongly to transitory shocks during 1994-1998.

Treisman 2005) or due to overstatement of inflation by the official CPI (Gibson *et al* 2004). To detect possible sources of CPI bias, we examine food prices and quantities from RLMS and find that the composition of food consumption has not changed much. Consequently, the CPI substitution bias within the food category is likely to be small. We do find, however, substantial disagreement in food CPI between RLMS and National Income and Product Accounts (NIPA), with a 25 percent discrepancy in the cumulative inflation during 1994-1998, but not much discrepancy afterwards. In addition, there is evidence of quality bias in the official CPI.

The paper's goal of documenting a comprehensive set of macroeconomic facts for Russia links it to many bodies of literature in macroeconomics, labor economics, development economics, and transition economics. In the interest of space, the literature survey below is necessarily incomplete, and it merely catalogues some of the related recent work by topic. Our analysis is closely related to the growing empirical literature that analyzes the joint evolution of income and consumption distributions (Cutler and Katz 1992, Attanasio and Davis 1996, Blundell and Preston 1998, Slesnick 2001, Krueger and Perri 2006, Heathcote *et al* 2007, Blundell *et al* 2008, etc.). There is also a growing body of research on inequality in developing countries. We find this literature particularly relevant for our study as it emphasizes the importance of measurement issues, urban-rural differences, home production, and income underreporting in understanding inequality in developing countries (e.g., Chen and Ravallion 1996, Deaton 1997).

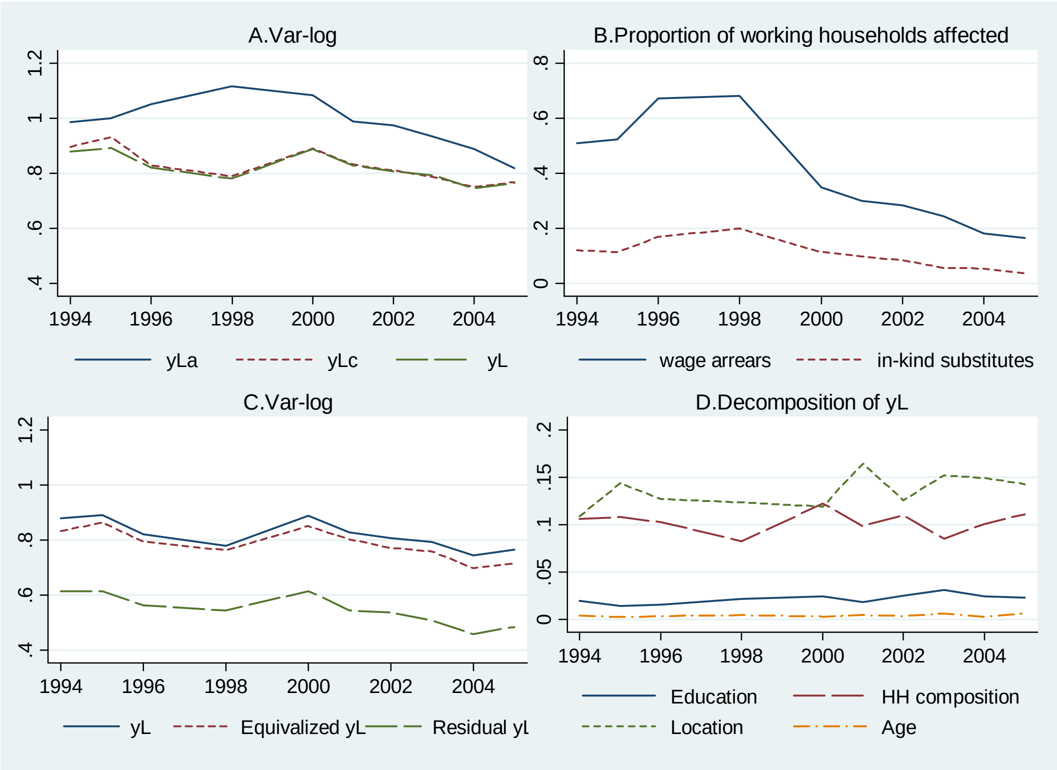
Several papers document changes in income inequality in Russia in the 1990s. These studies establish a number of important facts for the early transition period: rising income inequality, significant income mobility, large regional variation, and insufficient government transfers to offset an increase in wage inequality (Commander *et al* 1999; Milanovic 1999; Flemming and Micklewright 2000). The rise in income inequality is mainly attributed to compositional shifts from the old state sector to the new private sector, liberalization of wage setting, liberalization of prices and trade, and macroeconomic volatility. Some studies argue in favor of inequality measures based on expenditures (Aivazyan and Kolennikov 1999, Jovanovic 2001). They find a significant share of the transitory component in shocks to expenditures, high

Figure 6: Inequality in Labor Supply



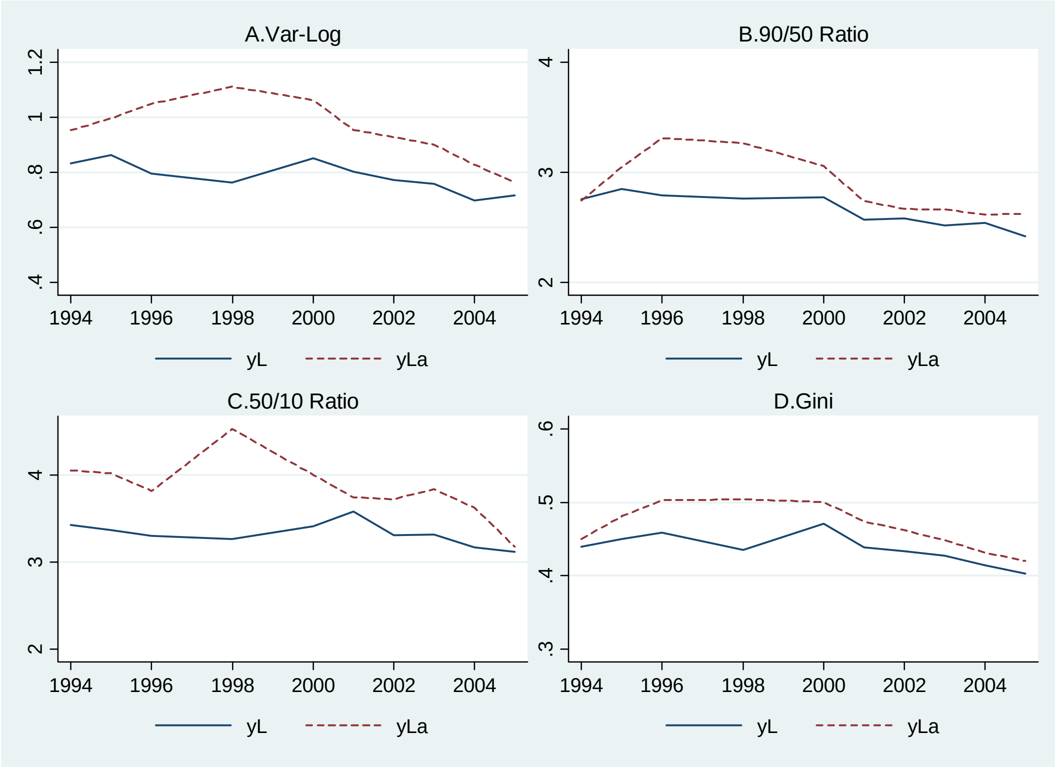
Notes: *wa* = hourly wage rate based on earnings received last month; *wc* = contractual hourly wage rate; *ha* = hours worked last month; *hc* = usual hours of work per month.

Figure 7: Household Earnings Inequality and Its Decomposition



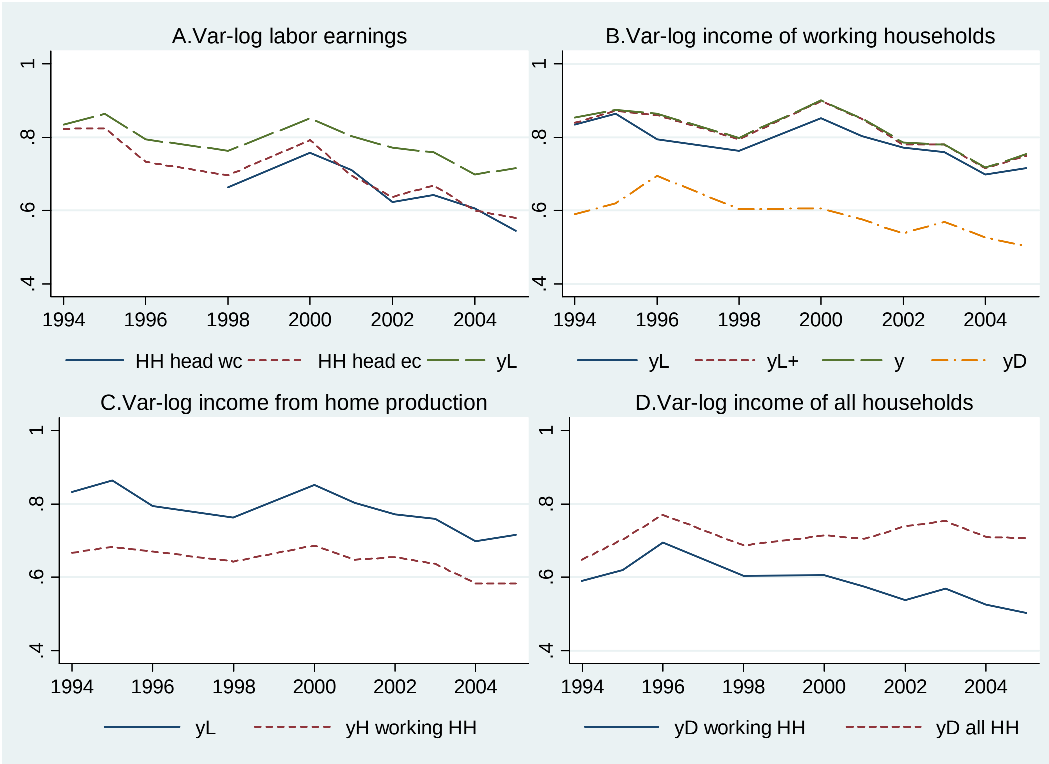
Notes: All earnings are after-tax and deflated using national monthly CPI. *yLa* = actual household labor earnings received last month; *yLc* = household contractual labor earnings per month; *yL* = household contractual labor earnings per month adjusted for non-response. Panel C reports the variance of log raw *yL*, the variance of log *yL* equivalized with an OECD equivalence scale, and the variance of residuals from equation (1). Panel D reports the variance of each observable component of equation (1).

Figure 8: Basic Inequality in Equivalized Household Earnings



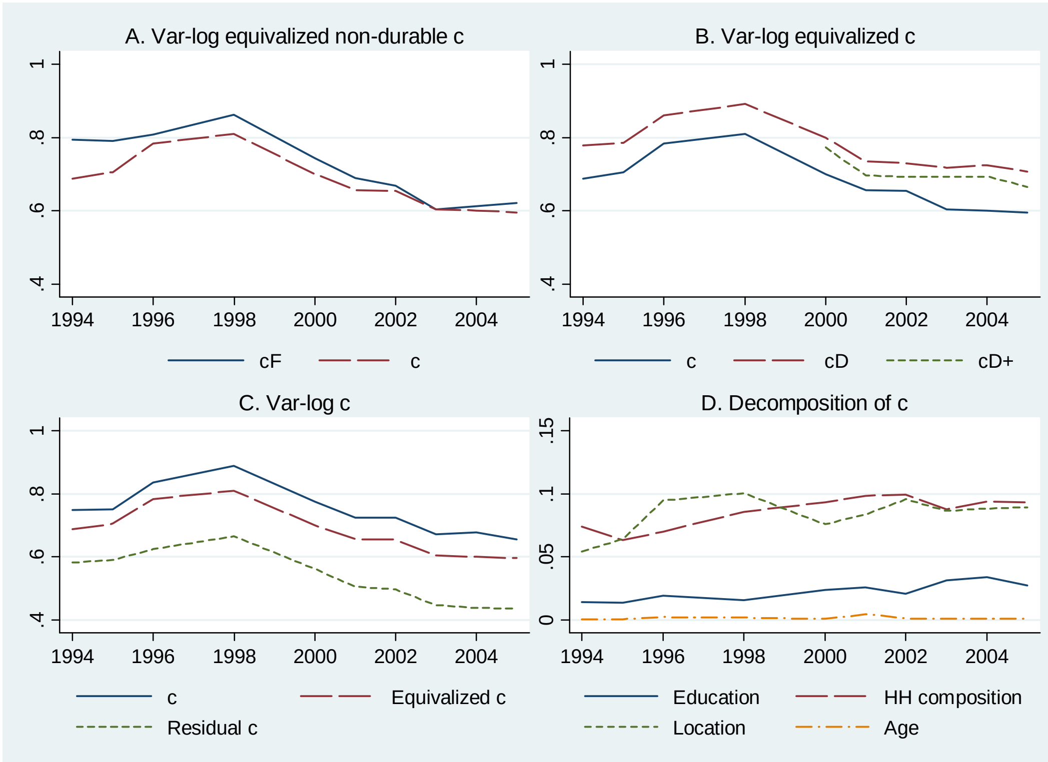
Notes: All earnings are after-tax, equivalized using an OECD equivalence scale, and deflated using national monthly CPI. *yLa* = actual household labor earnings received last month; *yL* = household contractual labor earnings per month adjusted for non-response.

Figure 9: From Wages to Disposable Income



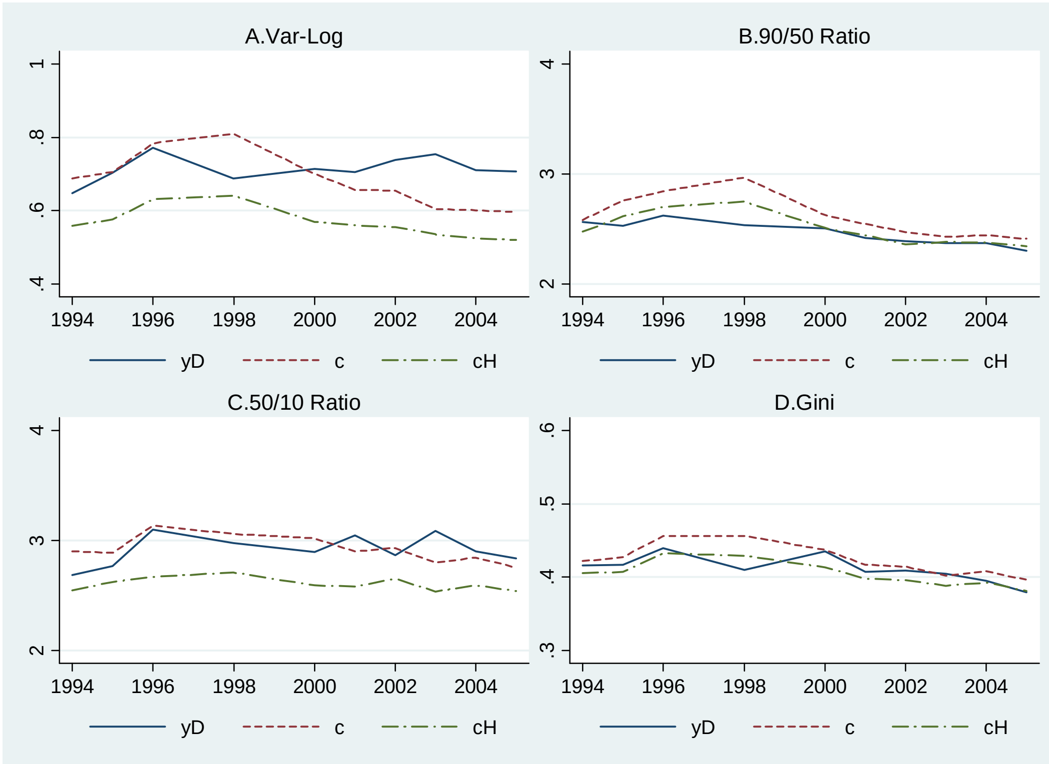
Notes: All income measures are after-tax and deflated using national monthly CPI. Measures at the household level are also equivalized using an OECD equivalence scale. *HH head wc* = contractual hourly wage rate of the head of household; *HH head ec* = contractual labor earnings per month of the head of household; *yL* = household contractual labor earnings per month adjusted for non-response; *yL+* = *yL+* private transfers; *y* = (*yL+*) + financial income; *yD* = disposable household income = *y* + government transfers; *yH* = *yL* + income from home production. Working households include households with at least one wage earner. Var-log is the variance of the logarithm of income.

Figure 10: Consumption Inequality and Its Decomposition



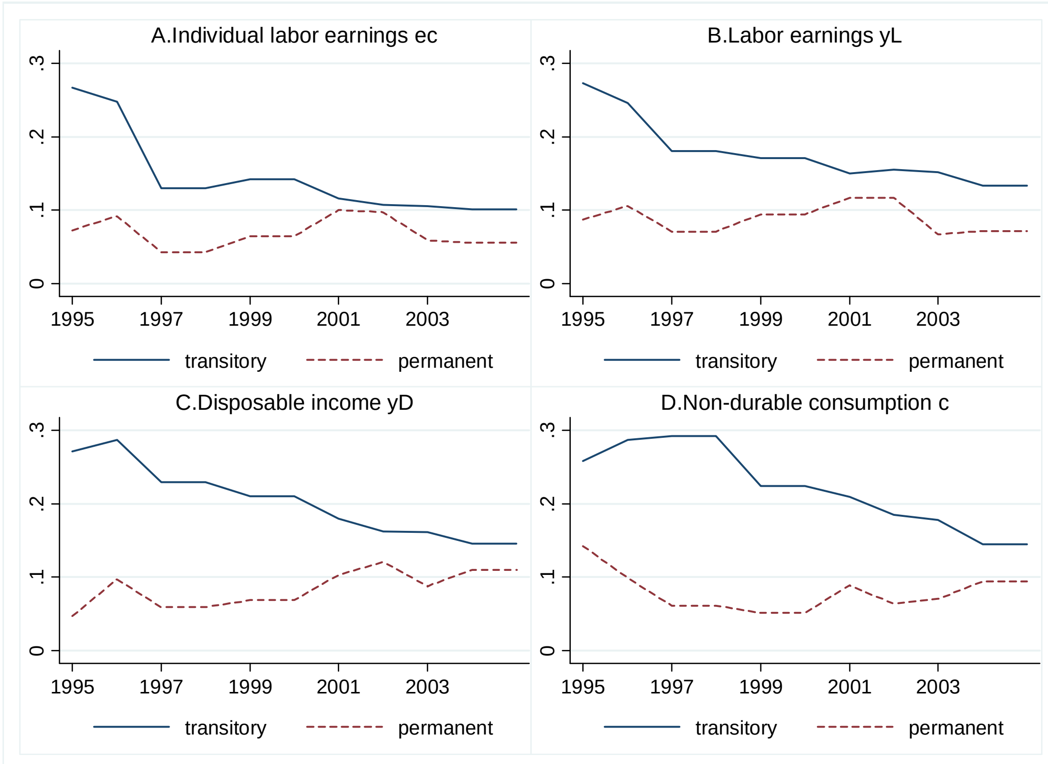
Notes: *cF* = expenditures on food, beverages, and tobacco last week (multiplied by 30/7); *c* = household non-durable expenditures last month; *cD* = *c* + expenditures on durables; *cD+* = *cD* + imputed services from housing. All consumption variables in Panels A and B are per adult equivalent. Panel C reports the variance of log raw *c*, the variance of log *c* equivalized with an OECD equivalence scale, and the variance of the residuals from equation (1). Panel D reports the variance of each observable component from equation (1).

Figure 11: From Disposable Income to Consumption



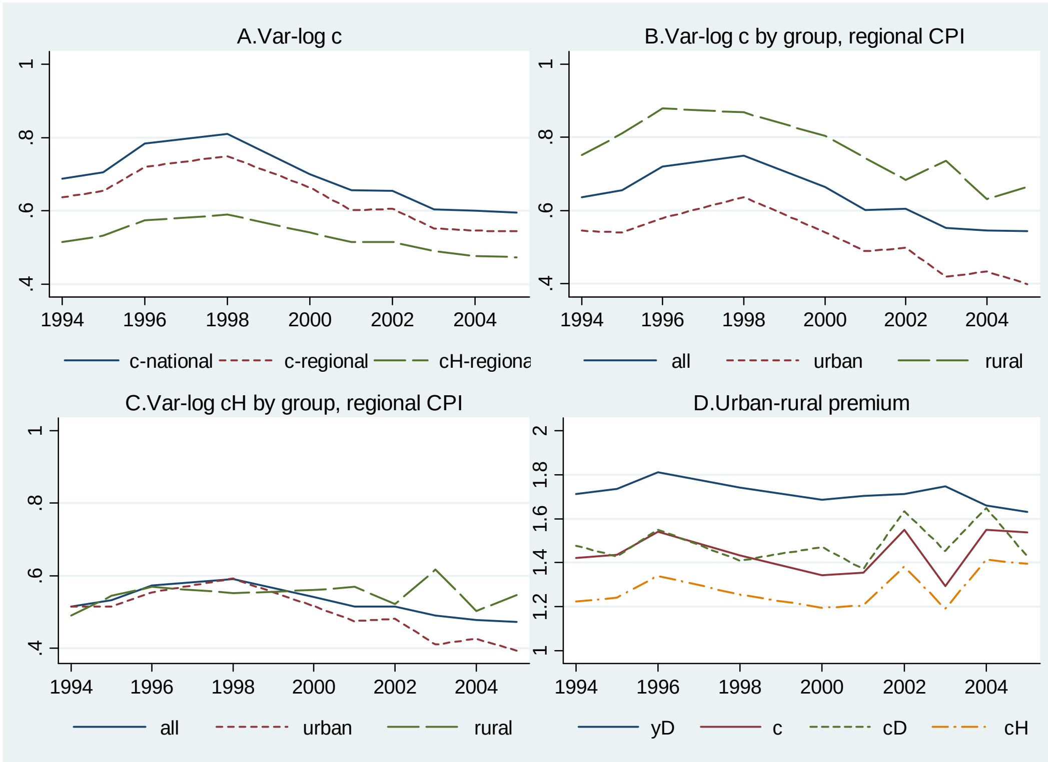
Notes: *yD* = disposable household income based on contractual labor earnings; *c* = household non-durable expenditures last month; *cH* = *c* + consumption of home-grown food. All measures are equivalized using an OECD equivalence scale and deflated with national monthly CPI.

Figure 12: Permanent-Temporary Component Decompositions



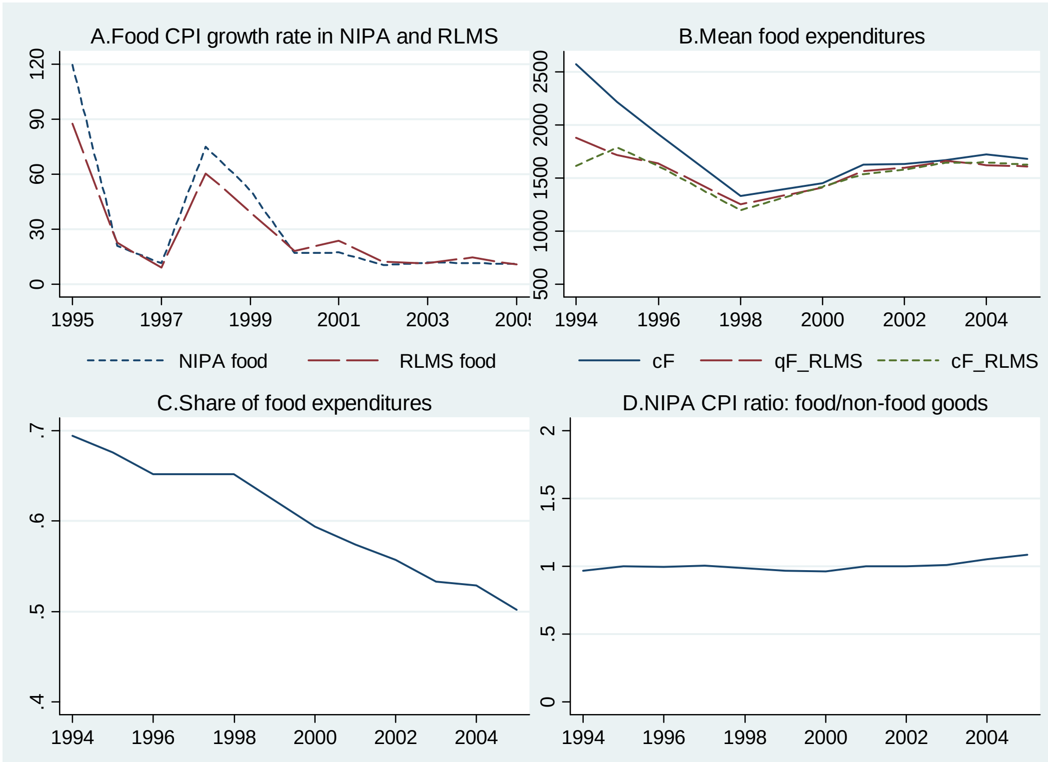
Notes: The figure reports the time series of estimated variance of permanent and transitory components. The estimated process is $u_{ht}=\alpha_{ht} + \varepsilon_{ht}$, $\alpha_{ht} = \alpha_{h,t-1} + \eta_{ht}$, where ε_{ht} is the transitory component and η_{ht} is the permanent component. In all specifications, u_{ht} is the residual from projecting the relevant measure of income or consumption on our baseline vector of observable characteristics of households; *ec* = contractual labor earnings of the household head; *yL* = household contractual labor earnings per month adjusted for non-response; *yD* = disposable household income based on contractual labor earnings; *c* = household non-durable expenditures last month. Values in 1998 and 2000 are adjusted for the fact that the permanent component is accumulated over two years. For both permanent and transitory components, 1997 and 1999 values are set equal to 1998 and 2000 values respectively.

Figure 13: Within-Group and Between-Group Inequality



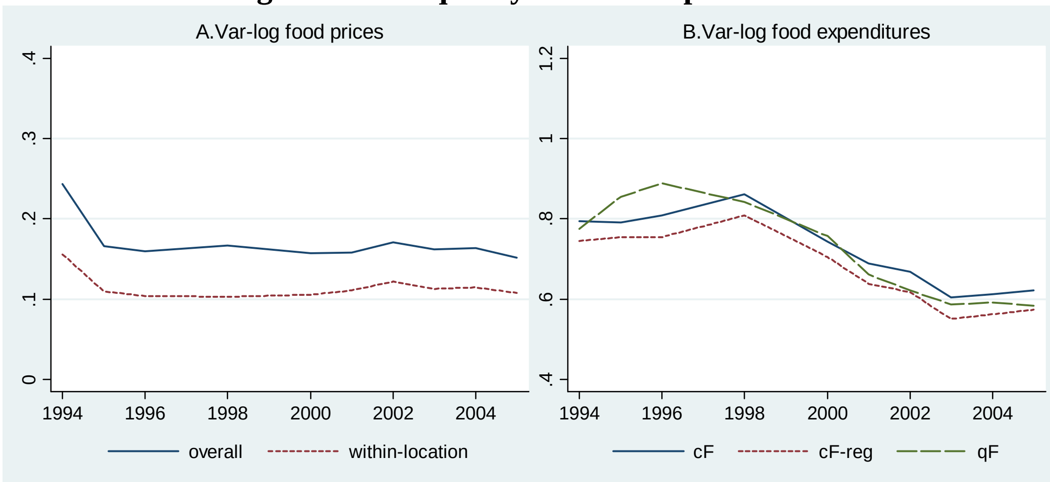
Notes: Rural location is defined as villages and small towns. *yD* = disposable household income based on contractual labor earnings; *c* = household non-durable expenditures last month; *cD* = *c* + expenditures on durables; *cH* = *c* + consumption of home-grown food. All measures are equivalized using an OECD equivalence scale and deflated with regional CPI unless indicated otherwise.

Figure 14: Trends in Food Expenditures



Notes: *cF_RLMS* = expenditures on food, beverages, and tobacco last week (multiplied by 30/7) deflated using national monthly CPI; *cF_RLMS* = expenditures on food, beverages, and tobacco last week deflated using RLMS food CPI; *qF_RLMS* = food quantity index in constant 2002 mean prices for each location. Panel C reports the share of food expenditures *cF* in aggregate consumption expenditures *cD*. All food expenditures are per adult equivalent.

Figure 15: Inequality in Food Expenditures



Notes: *cF* = expenditures on food, beverages, and tobacco last week (multiplied by 30/7) deflated using national monthly CPI; *cF-reg* = *cF* deflated using regional CPI and adjusted for regional differences in cost of living; *qF* = food quantity index in constant 2002 mean prices for each location. All food expenditures are per adult equivalent.

Appendix 1: Data Description

Description of RLMS sample

This study uses ten rounds of the Russian Longitudinal Monitoring Survey (RLMS) that was conducted in 1994-1996, 1998, and 2000-2005. RLMS was not conducted in 1997 and 1999. Time-series reported on the figures are linearly interpolated for missing annual data points.
The RLMS sample consists of the 38 randomly selected primary sample units (municipalities) that are representative of the whole country. They are located in 32 regions (or constituent subjects of the Russian Federation) and 7 federal districts. Russia had 89 constituent subjects and 7 federal districts as of December 1, 2005.

Sample restrictions

We restrict our sample to households in which at least one individual is 25-60 years old. The head of the household in the selected sample is the oldest working-age male or the oldest working-age female if no working-age males are present. If more than one person of the same age-gender is qualified for the head, then the reference person (or the first person surveyed in the roster files) is chosen.

General notes

1. All income variables are *after tax*.

2. All income and consumption variables are constructed on a *monthly basis*.

3. Summary statistics are weighted with individual and household sample weights provided in the RLMS.

4. When a household purchased the item but did not report the amount of the purchase, the missing amounts are imputed by regressing the log of expenditure on the complete interaction between year dummies and federal district dummies, controlling for the size of the household (5 categories), number of children 16 years old or younger (4 categories), number of elderly members 60+ (3 categories), and urban location. Because of the log dependent variable, the predicted values of expenditures are adjusted as $y = exp(\hat{\sigma}^2/2)exp\left(\widehat{\log y}\right)$. The subcategories with the largest number of missing values include utilities (2.12% of the sample), gasoline and motor oil (1.63%), transportation services (1.54%), and contributions to non-relatives (1.35%). Missing values for other subcategories are trivial.

5. Similar regression-based imputations are performed for missing subcategories of non-labor income and income from home production. Imputations of labor income are described in the table below. Although the share of missing values for each individual subcategory of non-labor income and expenditures is very small, altogether missing values affect about a third of surveyed households. Our imputation procedure is an improvement over the existing RLMS practice that treats missing values as zeros in computing aggregate income and expenditures.

Variable description and notes

	Variable Name	Definition	Notes
		Individual Earnings and Labor Supply	
ha	Actual hours of work last month	= hours worked last month at primary job + hours worked last month at secondary job + hours spent last month on regular individual economic activities (activities for which an individual is paid for regularly, such as sewing a dress, assisting with repairs, selling goods in a market or on the street, etc.)	Unusually high hours are top coded at 480 hours per month (16 hours per day*30 days)
hc	Usual hours of work per month	= 4 times usual hours in a typical week at primary job + 4 times usual hours in a typical week at secondary job + hours spent last month on regular individual economic activities.	*hc* is available in 1998-2005 only. Unusually high hours are top coded at 480 hours per month (16 hours per day*30 days).
status	Working status	= *full-time* if actual hours at primary job≥120, *part-time* if actual hours at primary job<120, *not working* if a respondent did not work last month at primary job, was not on a temporary leave, and was not engaged in regular individual economic activities	
ea	Actual labor earnings last month	= money received last month from primary job + money received last month from secondary job +	The variable is highly volatile during the period of wage arrears since a worker may

		money received last month from regular individual economic activities + payments in kind received last month from primary job + payments in kind received last month from secondary job	not receive any money last month or receive back payments for several months at once.
ec	Contractual labor earnings per month	*1998-2005, all employees*: = monthly average (over the last 12 months) after-tax labor earnings of an employee at primary job + money received last month from additional jobs for all employees in 1998-2005 *1994-1996, employees with wage arrears*: = total accumulated wage debt divided by the number of months of overdue wages + money received last month from additional jobs for employees with wage arrears at primary job in 1994-1996 *1994-1996, employees with no wage arrears*: = monetary portion of *wa* for employees with no wage arrears *All years, self-employed*: = monetary portion of *wa* for self-employed (or individuals reporting place of work other than an organization), including those involved in regular individual economic activities in all years.	1. *ec* does not include payments in kind. 2. Average monthly earnings are available for an employee at primary job in 1998-2005. 3. Implausibly low earnings below ½ of the official minimum monthly wage are recoded into missing (0.47% of positive earnings). 4. Implausibly high earnings are also recorded into missing if the residuals exceed five standard deviations from the mean after controlling for occupational categories, hours of work, age, age squared, years of schooling, and individual fixed effects (0.13% of positive earnings). 5. For household aggregation purposes, if a respondent worked last month at least one hour but has missing contractual earnings, missing values are imputed using occupational categories, hours of work, gender, age, age squared, years of schooling, urban location and federal district dummies (the share of imputed earnings is 7.8%).
wa	Hourly wage rate last	= *ea* / *ha*	

	month		
wc	Contractual hourly wage rate	= *ec* / *hc*	*hc* is available in 1998-2005 only; *wc* is calculated for non-imputed earnings
		Household Income	
yLa	Actual labor earnings received last month	After-tax payments received by all household members from all places of work in the form of money, goods, and services in the last 30 days as reported by the reference person of the household.	The variable is highly volatile during the period of wage arrears.
yLc	Contractual labor earnings per month	The sum of *ec* across all individual respondents within the household.	Such aggregation omits those adult household members who did not respond to an individual questionnaire; the response rate for working age individuals within the surveyed household is 96.5%.
yL	Contractual labor earnings per month adjusted for non-response	= *yLc* + imputed contractual labor earnings for working-age non-respondents within the household.	Labor earnings of working-age non-respondents are imputed as predicted earnings times the predicted probability of working using the full set of interactions between the four age groups (18-60) and two gender groups and controlling for urban and federal district dummies for each year separately.
yH	Labor earnings plus income from home production	= *yL* + 0.9*h*, where *h* is average monthly income from home-grown food in the last year defined as the sum of physical quantity of produced food items (minus items given away) multiplied by their mean price in a given region, 0.9 is the assumed labor share of home food production.	Mean prices are obtained in two steps. First, the household-specific market price of individual food item is calculated by dividing the cost of purchase by the amount purchased in the last 7 days. Then the mean price of individual food items is computed for each region (*oblast*) and year.

yL+	Labor earnings plus net private transfers	= *yL* + private transfers received last month – private transfers given to individuals outside the household unit last month.	"Private transfers received" include received alimonies and 11 subcategories of contributions from persons outside the household unit, including contributions from relatives, friends, charity, international organizations, etc. "Private transfers given" include alimonies paid and various contributions in money and in kind given to individuals outside the household unit (6 categories).
y	Household income before government transfers	= *yL* + net private transfers + financial income received last month.	Financial income includes dividends on stocks and interest on bank accounts.
yD	Disposable household income	= *y* + public transfers.	Public transfers include government pensions, state child benefits, stipends, unemployment benefits, and government welfare payments.
		Household Consumption	
cF	Market expenditures on food, alcohol and tobacco	Monthly expenditures on food, alcohol, and tobacco are computed as the sum of expenditures on individual items in the reference week multiplied by 30/7=4.286.	Items include 50 categories of food at home and away from home, alcoholic and non-alcoholic beverages, and tobacco products. See Appendix 2 for details of computation.
qF	Food quantity index	$qF_t = \sum_k \overline{p}_k q_{kt}$, where q_{kt} is the quantity of food item *k* purchased in year *t* and $\overline{p}_k$ is average price of item *k* for each location (psu) in the base year (2002).	
c	Non-durable expenditures	Sum of expenditures on non-durables in the last 30 days. Non-durable items include food, alcohol, tobacco, clothing and footwear, gasoline	

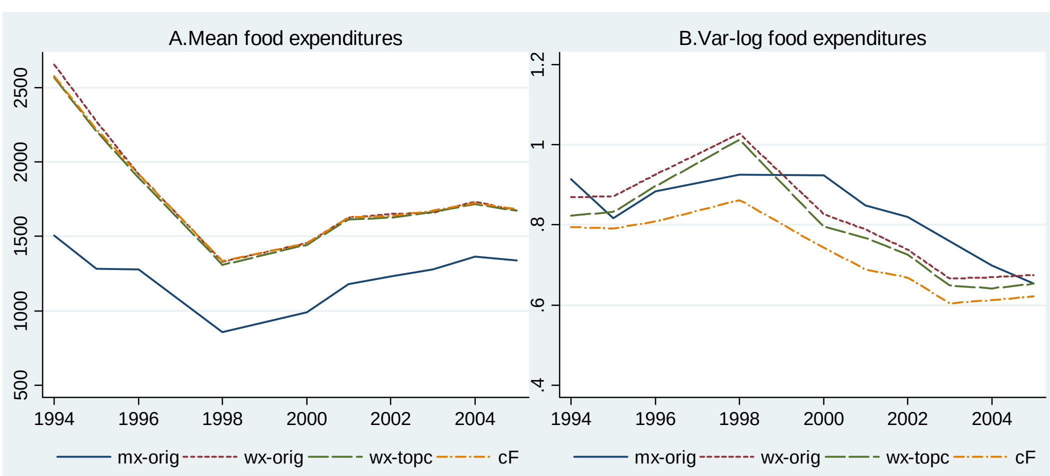
		and other fuel expenses, rents and utilities, and 15-20 subcategories of services (such as transportation, repair, health care services, education, entertainment, recreation, insurance, etc.).	
cD	Aggregate expenditures	= *c* + expenditures on durables in the last 3 months / 3. Durable items include 10 subcategories such as major appliances, vehicles, furniture, entertainment equipment, etc.	This is compared with purchases of goods and services from NIPA
cH	Non-durable expenditures plus consumption of home-grown food	= *c* + consumption of home-grown food, where the last term is calculated as average monthly quantities of consumed home-grown food items multiplied by their mean price in a given region.	Mean prices are determined in the same way as in *yH*.
cD+	Aggregate expenditures plus services from housing	= *cD* + imputed services from housing.	Imputed services from housing are calculated as 5% of the current housing market value divided by 12.
		Adjustments to Income and Consumption	
equiv	OECD equivalence scale	This equivalence scale assigns a value of 1.0 to the first adult household member, a value of 0.7 to each additional adult, and a value of 0.5 to each child 16 years old and younger.	
cpi_t	National monthly CPI	All income and consumption variables are deflated in prices of 2002 using monthly national CPI.	If the date of interview is in the first half of month, the previous month CPI is used. If the date of interview is in the second half of month, the current month CPI is used.
def_t	Regional deflator	Deflator that combines monthly national CPI, December to December regional CPIs, and the	To adjust for monthly inflation, the flow variables are expressed in December prices of

		regional value of fixed basket of goods and services.	each year by using a country average monthly CPI and the date of interview. Next, the annual (December to December) CPI for each 32 oblasts (regions) is applied to convert the flow variables into prices of December 2002. Finally, these real values are adjusted for regional differences in the cost-of-living by using the regional value of fixed basket of goods and services.
$cpiF_{RLMS,t}$	RLMS food CPI	$cpiF_{RLMS,t} = \sum_k p_{kt}\bar{q}_k \Big/ \sum_k \bar{p}_k \bar{q}_k$, where p_{kt} denote the sample average unit price of food category k in year t; $\bar{p}_k$ and $\bar{q}_k$ are the sample average price and the quantity of food item k purchased in the base year.	
		Control Variables	
D^H	Household composition	Vector of household composition variables: 4 categories for the number of children 16 years old and younger (0, 1, 2, and 3+), 3 categories for the number of seniors 60 years old and older (0, 1, and 2+), and 5 categories for the number of household members (1, 2, 3, 4, and 5+).	
	Demographics	A female dummy and continuous age variable a.	
D^E	Schooling	A set of dummies for educational attainment of the head of household (incomplete secondary, secondary, vocational, technical, and university)	
D^L	Location variables	A set of dummies for 7 federal districts, a dummy for Moscow and St. Petersburg, and a dummy for urban location.	

Appendix 2: Constructing Food Expenditures

This appendix describes the steps in constructing our measure of food expenditures.

1. RLMS food data contain information on the physical quantity and monetary value of last week purchases for 50 categories of food at home and away from home, alcoholic and non-alcoholic beverages, and tobacco products. We first create *wx-orig* as the sum of expenditures on these individual items multiplied by 30/7=4.286. Missing values for this measure are treated as zero.
2. The RLMS questionnaire also asks about the total sum of food purchases in the last 30 days (*mx-orig*). We discard this measure because of a potentially large measurement error, higher probability of underreporting, and ambiguity in the question (e.g., it is likely to exclude beverages and tobacco). We note, however, that the two measures of food expenditures have similar variance (compare *wx-orig* and *mx-orig* in figure below).
3. When a household purchased the item but did not report the quantity of the purchase, the missing quantities are imputed by regressing the log of expenditure on the complete interaction between year dummies and federal district dummies, controlling for the size of the household (5 categories), number of children 16 years old or younger (4 categories), number of elderly members 60+ (3 categories), and urban location. Because of the log dependent variable, the predicted values of expenditures are adjusted as $y = exp(\hat{\sigma}^2/2)exp(\widehat{\log y})$. Missing values for food items are generally trivial.
4. We use top coding of unreasonably high prices in excess of 3 interquantile ranges above the mean prices in a given location as well as unreasonably high amounts (quantities) of food purchases (the top 99th percentile), conditional on the household structure and location. Top coding and imputations does not change the mean value and only slightly reduce the variance (see *wx-topc* in figure below)
5. It is very well known that inequality measures, especially those based on logarithms, are very sensitive to very low values. For that reason, we eliminate the bottom 1% of total food consumption (from purchases and home production) in constant 2002 prices (about 12 percent of the cost of the reference basket of 25 major food items reported by Goskomstat in 2002). While this procedure does not change the mean value of food expenditures, it predictably reduces the variance (see line *cF*).



Notes: All reported measures are per adult equivalent and deflated with national monthly CPI.

Appendix 3: Sample Composition

		Full sample	Restricted sample	Estimation sample
Year:	1994	9.34	9.61	9.66
	1995	8.89	9.09	9.07
	1996	8.82	8.94	8.75
	1998	9.00	9.01	8.91
	2000	9.42	9.24	9.23
	2001	10.64	10.35	10.42
	2002	10.97	10.74	10.81
	2003	11.09	10.92	10.96
	2004	11.07	11.17	11.21
	2005	10.75	10.92	10.99
Region:	Moscow and St. Petersburg	11.28	11.17	11.31
	North West	6.89	7.33	7.37
	Central	19.09	18.17	18.26
	Volga	17.72	17.42	17.39
	South	11.73	12.13	11.93
	Urals	14.17	14.60	14.59
	Siberia	9.41	9.45	9.41
	Far East	9.71	9.73	9.73
Number of household members: 1		18.39	7.58	7.18
	2	27.74	24.28	24.16
	3	25.34	30.83	31.07
	4	18.06	23.49	23.72
	5+	10.47	13.82	13.87
Number of children <16: None		56.99	45.63	45.25
	1	28.26	35.02	35.32
	2	12.23	15.99	16.09
	3+	2.53	3.36	3.34
Urban (excluding small towns)		68.91	69.55	70.01
		42,541	31,969	31,409

Notes: Restricted sample includes households in which at least one individual is 25-60 years old. Estimation sample includes households with non-missing values on disposable income. The sample composition is unweighted.

Appendix 4: Inequality over the Life Cycle

The peculiar age-earnings profile for Russian males with its negative experience premium (Figure 5C) underscores the importance of investigating the life-cycle pattern of inequality. One would like to separate out the age effect on inequality from time effects and cohort effects but these effects are collinear unless one imposes additional restrictions (see Heathcote *et al* 2008). Since none of the restrictions is entirely satisfactory, we present decompositions of age, cohort, and time effects under alternative identifying assumptions.

Suppose that the cross-sectional inequality moment $M(a,t)$ depends on age, a, time, t, and cohort effects, $t-a$, through a linear function. An inequality-age regression can separately identify one of these three effects, and the combined effect of the other two. We first perform inequality-age regressions controlling for time effects and assuming that there are no cohort effects. This specification confounds age effects with cohort effects, and the two cannot be separately identified. We regress the inequality moments on the set of age and time dummies:

$$M(a,t) = \sum_a \beta_a D(a) + \sum_t \beta_t D(t) + \varepsilon_{a,t},$$

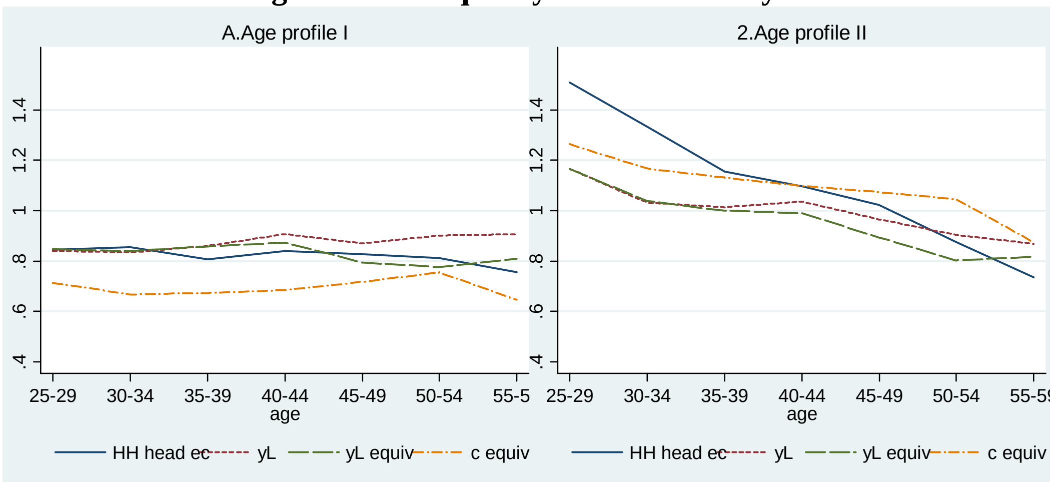
where $M(a,t)$ denote the variance of log income (or consumption) within age group a at time t. Panel A below shows the pattern of age dummies β_a. In almost all cases, the age-inequality profiles are essentially flat, with the exception of a slight decline in inequality among the oldest workers. The flat life cycle inequality profile can be interpreted as age effects and cohort effects roughly canceling each other out. The flat profile of age dummies is consistent with income and consumption decompositions in Figures 7 and 10, where age was found to have almost no explanatory power.

Panel B below reports the age coefficients β'_a from a different specification that assumes away time effects and regresses the cross-sectional inequality moments on age and cohort dummies:

$$M(a,t) = \sum_a \beta'_a D(a) + \sum_{t-a} \beta'_{t-a} D(t-a) + \varepsilon'_{a,t-a}.$$

Now the age-inequality profiles are downward-sloping, because time effects are confounded with age effects. In other words, if income and consumption inequality falls over time for a fixed cohort, the regression model categorizes this as an age effect. Our results potentially point to large time effects on inequality.

Figure A1: Inequality over the Life Cycle



Notes: Panel A depicts age profiles for the var-log controlling for year effects. Panel B depicts age profiles for the var-log controlling for cohort effects. All measures are deflated with national monthly CPI. *HH head ec* = contractual labor earnings per month of the head of household; *yL* = household contractual labor earnings per month adjusted for non-response; *yL equiv* = *yL* equivalized with an OECD equivalence scale; *c equiv* = household non-durable expenditures equivalized with an OECD equivalence scale.

Appendix 5: Time Series Decomposition of Income and Consumption Growth

Permanent-temporary decomposition

The procedure decomposes *residual* variation of income and consumption $u_{ht}^{(s)}$ into temporary and permanent components, where s denotes a measure of income or consumption. Using the notation in the body of the paper, the residual $u_{ht}^{(s)}$ from regression (1) can be decomposed into the sum of a transitory component and a random-walk permanent component:

$$u_{ht}^{(s)} = \alpha_{ht} + \varepsilon_{ht},$$

$$\alpha_{ht} = \alpha_{h,t-1} + \eta_{ht},$$

where $\varepsilon_{ht} \sim (0, \sigma_{\varepsilon,t}^2)$ is the transitory component and $\eta_{ht} \sim (0, \sigma_{\eta,t}^2)$ is the innovation in the permanent component.

Given $u_{ht}^{(s)}$, we form a vector of changes in the residual $\Delta u_{ht}^{(s)} = \eta_{ht} + \varepsilon_{ht} - \varepsilon_{h,t-1}$ (that equals the annual growth rate of s_{ht}). The full vector of growth rates for household h and measure s_{ht} is $g_h = [\Delta u_{h,1}^{(s)}\ \Delta u_{h,2}^{(s)}\ ...\Delta u_{h,T}^{(s)}]'$, where t = 0 is the first year in the panel and T is the last. The covariance matrix of vector g_h, which has $T(T-1)/2$ unique empirical moments, is

$$V \equiv \tfrac{1}{H}\sum_{h=1}^{H}(g_h - \overline{g})(g_h - \overline{g})'$$

where $\overline{g} = \tfrac{1}{H}\sum_{h=1}^{H} g_h$ is the average value of the change in the residual and H is the number of households in the sample.

Let Λ be the vector of parameters we to be estimated (i.e., the year-specific variances of innovations in permanent and transitory components of s_{ht}) and let $V(\Lambda)$ be the corresponding covariance matrix. Under the assumptions of our statistical model,

$$V(\Lambda) = \begin{bmatrix} \sigma_{\eta,1}^2 + \sigma_{\varepsilon,1}^2 + \sigma_{\varepsilon,0}^2 & -\sigma_{\varepsilon,1}^2 & 0 & \cdots & 0 & 0 \\ -\sigma_{\varepsilon,1}^2 & \sigma_{\eta,2}^2 + \sigma_{\varepsilon,2}^2 + \sigma_{\varepsilon,1}^2 & -\sigma_{\varepsilon,2}^2 & \ddots & 0 & 0 \\ 0 & -\sigma_{\varepsilon,2}^2 & \sigma_{\eta,3}^2 + \sigma_{\varepsilon,3}^2 + \sigma_{\varepsilon,2}^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \sigma_{\eta,T-1}^2 + \sigma_{\varepsilon,T-1}^2 + \sigma_{\varepsilon,T-2}^2 & -\sigma_{\varepsilon,T-1}^2 \\ 0 & 0 & 0 & \ddots & -\sigma_{\varepsilon,T-1}^2 & \sigma_{\eta,T}^2 + \sigma_{\varepsilon,T}^2 + \sigma_{\varepsilon,T-1}^2 \end{bmatrix}.$$

Two identification issues are apparent from the above expression for $V(\Lambda)$. First, $\sigma_{\varepsilon,0}^2$ is not identified separately from $\sigma_{\eta,1}^2$. Second, $\sigma_{\varepsilon,T}^2$ is not identified separately from $\sigma_{\eta,T}^2$. We follow the common practice of addressing these identification issues by imposing $\sigma_{\varepsilon,T}^2 = \sigma_{\varepsilon,T-1}^2$ and $\sigma_{\varepsilon,1}^2 = \sigma_{\varepsilon,0}^2$. After imposing these constraints, the vector of parameters to be estimated becomes $\Lambda = \{\sigma_{\varepsilon,1}^2, \sigma_{\varepsilon,2}^2, ..., \sigma_{\varepsilon,T-1}^2, \sigma_{\eta,1}^2, \sigma_{\eta,2}^2, ..., \sigma_{\eta,T}^2\}$.

Vector Λ is estimated by minimizing the distance between theoretical and empirical moments

$$\hat{\Lambda} = \arg\max_{\Lambda}(vech\{V - V(\Lambda)\})'(vech\{V - V(\Lambda)\}),$$

where the weight matrix is set to be the identity matrix.

Estimating consumption response to income innovations

The approach to estimating the response of consumption to income components is similar to that in Blundell *et al* (2008). The procedure uses (auto)covariances of income and consumption growth rates. As before, the residual in the income equation is assumed to follow the process

$$u_{ht}^{(y)} = \alpha_{ht} + \varepsilon_{ht},$$
$$\alpha_{ht} = \alpha_{h,t-1} + \eta_{ht},$$

The residual consumption growth

$$\Delta u_{ht}^{(c)} = \phi_t \eta_{ht} + \psi_t \varepsilon_{ht} + \xi_{ht} - \xi_{h,t-1},$$

is decomposed into the parts: the influence of permanent income innovation, the influence of temporary income innovation, and unobserved household heterogeneity. Let $g_h = \left[\Delta u_{h,1}^{(c)}, \Delta u_{h,1}^{(y)}, ..., \Delta u_{h,T}^{(c)}, \Delta u_{h,T}^{(y)}\right]$ denote the vector of income and consumption growth rates for household h. As before, define the empirical covariance matrix

$$V \equiv \tfrac{1}{H}\sum_{h=1}^{H}(g_h - \overline{g})(g_h - \overline{g})'.$$

Let Λ be the vector of parameters we to be estimated (i.e., the year-specific variances of innovations in permanent and transitory components of income and transitory components in consumption as well as loadings ϕ_t and ψ_t) and let $V(\Lambda)$ be the vector of theoretical moments (i.e., the model equivalent of V). Under our statistical model, with T=3 (for example) we have

$$V(\Lambda) = \begin{bmatrix} \sigma_{\eta,1}^2 + \sigma_{\varepsilon,1}^2 + \sigma_{\varepsilon,0}^2 & \phi_1^2\sigma_{\eta,1}^2 + \psi_1^2\sigma_{\varepsilon,1}^2 & -\sigma_{\varepsilon,1}^2 & 0 & 0 & 0 \\ \phi_1^2\sigma_{\eta,1}^2 + \psi_1^2\sigma_{\varepsilon,1}^2 & \phi_1^2\sigma_{\eta,1}^2 + \psi_1^2\sigma_{\varepsilon,1}^2 + \sigma_{\xi,1}^2 + \sigma_{\xi,0}^2 & -\psi_1\sigma_{\varepsilon,1}^2 & -\sigma_{\xi,1}^2 & 0 & 0 \\ -\sigma_{\varepsilon,1}^2 & -\psi_1\sigma_{\varepsilon,1}^2 & \sigma_{\eta,2}^2 + \sigma_{\varepsilon,2}^2 + \sigma_{\varepsilon,1}^2 & \phi_2^2\sigma_{\eta,2}^2 + \psi_2^2\sigma_{\varepsilon,2}^2 & -\sigma_{\varepsilon,2}^2 & 0 \\ 0 & -\sigma_{\xi,1}^2 & \phi_2^2\sigma_{\eta,2}^2 + \psi_2^2\sigma_{\varepsilon,2}^2 & \phi_2^2\sigma_{\eta,2}^2 + \psi_2^2\sigma_{\varepsilon,2}^2 + \sigma_{\xi,2}^2 + \sigma_{\xi,1}^2 & -\psi_2\sigma_{\varepsilon,2}^2 & -\sigma_{\xi,2}^2 \\ 0 & 0 & -\sigma_{\varepsilon,2}^2 & -\psi_2\sigma_{\varepsilon,2}^2 & \sigma_{\eta,3}^2 + \sigma_{\varepsilon,3}^2 + \sigma_{\varepsilon,2}^2 & \phi_3^2\sigma_{\eta,3}^2 + \psi_3^2\sigma_{\varepsilon,3}^2 \\ 0 & 0 & 0 & -\sigma_{\xi,2}^2 & \phi_3^2\sigma_{\eta,3}^2 + \psi_3^2\sigma_{\varepsilon,3}^2 & \phi_3^2\sigma_{\eta,3}^2 + \psi_3^2\sigma_{\varepsilon,3}^2 + \sigma_{\xi,3}^2 + \sigma_{\xi,2}^2 \end{bmatrix}.$$

Again, there are two identification issues. First, $\sigma_{\varepsilon,0}^2, \sigma_{\xi,0}^2$ are not identified separately from $\sigma_{\eta,1}^2$. Second, $\sigma_{\varepsilon,T}^2, \sigma_{\xi,T}^2$ are not identified separately from $\sigma_{\eta,T}^2$. We impose $\sigma_{\varepsilon,T}^2 = \sigma_{\varepsilon,T-1}^2, \sigma_{\varepsilon,1}^2 = \sigma_{\varepsilon,0}^2$, $\sigma_{\xi,T}^2 = \sigma_{\xi,T-1}^2, \sigma_{\xi,1}^2 = \sigma_{\xi,0}^2$. Thus, our vector of parameters becomes

$$\Lambda = \{\sigma_{\varepsilon,1}^2, \sigma_{\varepsilon,2}^2, ..., \sigma_{\varepsilon,T-1}^2, \sigma_{\eta,1}^2, \sigma_{\eta,2}^2, ..., \sigma_{\eta,T}^2, \sigma_{\xi,1}^2, \sigma_{\xi,2}^2, ..., \sigma_{\xi,T-1}^2, \phi_1, \phi_2, ..., \phi_T, \psi_1, \psi_2, ..., \psi_T\}.$$

We estimate Λ by minimizing the distance between theoretical and empirical moments

$$\hat{\Lambda} = \arg\max_{\Lambda}(vech\{V - V(\Lambda)\})'(vech\{V - V(\Lambda)\}),$$

where the weight matrix is set to be the identity matrix.